JOURNAL OF BACTERIOLOGY, Dec. 1987, p. 5633–5640
0021-9193/87/125633-08$02.00/0

Vol. 169, No. 12

Streptococcus pyogenes Type 12 M Protein Gene Regulation by Upstream Sequences

JOHN C. ROBBINS, JONATHAN G. SPANIER,† S. J. JONES, WARREN J. SIMPSON, AND P. PATRICK CLEARY*

Department of Microbiology, University of Minnesota, Minneapolis, Minnesota 55455

Received 9 September 1987/Accepted 17 September 1987

A partial nucleotide sequence that included 1,693 base pairs of the M12 (*emm12*) gene of group A streptococci (strain CS24) and adjacent upstream DNA was determined. Type 12 M protein-specific mRNA of strain CS24 is transcribed from two promoters (P_1 and P_3) separated by 30 bases. The transcription start sites of the *emm12* gene were located more than 400 bases downstream of a deletion that causes decreased M-protein gene transcription in strain CS64. Deletion analysis of M protein-expressing plasmids indicated that an upstream region greater than 1 kilobase is required for M-protein gene expression. The M-protein gene transcriptional unit appears to be monocistronic. Analysis of the *emm12* DNA sequence revealed three major repeat regions. Two copies of each repeat, A and B, existed within the variable 5′ end of the gene; repeat C demarcated the 5′ end of the constant region shared by *emm12* and *emm6*.

M protein protects *Streptococcus pyogenes* from phagocytosis by human polymorphonuclear leukocytes (20, 23). The change by group A streptococci from high-level expression of this surface protein (M^+) to a nonexpressive (M^-) state was shown by Todd and Lancefield (42) to parallel the change from an opaque (matt) to a transparent (glossy) colony phenotype; in a more recent report (35) it has been shown that the switch from high to low levels of M-protein expression occurs at high frequency and is reversible, which are characteristics of phase variation described for other bacterial species (7, 26). In addition to phase variation of M-protein expression (35), streptococci also exhibit both antigenic variation (more than 70 strains with antigenically distinct M proteins have been described) and size variation of this protein (11). The mechanisms that control phase and antigenic variation and the relationship between them is not understood. Although the nature of M protein as an antiphagocytic molecule has been demonstrated, a role for the M^- state in the survival of the organism has not been identified. Phase variation has been postulated to be a virulence factor that controls the expression of the type 1 pilus of *Escherichia coli* (6, 7) and the pilus of *Neisseria gonorrhoeae* (26).

The phase switch in M-protein expression by the M^+ group A streptococcal strain CS24 can produce nonrevertible, phase-locked variants which have suppressed levels of type 12 M protein (35). In one such variant, strain CS64, this change is accompanied by a deletion of approximately 50 base pairs (bp) near the 5′ end of the structural gene (39). Recently, it was shown that phase variation of type 12 M protein can occur concomitantly with either gross genomic rearrangements neighboring the M 12 gene (deletions detected by Southern analysis) or by alterations that have not been identified by Southern analysis. The former case represents a class of phase variants that are phase locked; i.e., they do not revert to the M^+ state; on the other hand, the latter type of phase variants is characterized by a revertible phenotype, switching at a high frequency (35). Here we report results of a study of the role of upstream neighboring sequences in the regulation of expression of the *emm12* gene of strain CS24 in *E. coli*, describe the physical relationship between the deletion in strain CS64 and the 5′ end of the *emm12* gene, and study the transcriptional activity of the *emm12* gene by Northern and primer extension analyses. We found that transcription of the *emm12* gene in strain CS64 is diminished by more than 100-fold and that the deletion is more than 400 bases upstream from the transcription start sites of the *emm12* gene.

MATERIALS AND METHODS

Bacterial strains, plasmids, bacteriophage, and media. M type 12 group A streptococcal strain CS24 and variant CS64 have been described previously (3, 39). Liquid cultures of group A streptococci were grown in Todd-Hewitt broth supplemented with 1% Neopeptone (Difco Laboratories, Detroit, Mich.) for 15 to 18 h at 37°C. *E. coli* JM83, JM103, and JM110, carrying constructs of plasmids pUC19 and pUC18, were propagated in L broth or on L agar containing ampicillin (30 μg/ml) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (0.03%) at 37°C. All plasmids used and constructed in this study are listed in Fig. 1.

Plasmids pPC101, pPC106, and pPC113 have been described previously (39). Plasmid pPC134 is a subclone of lambda clone EMBL3-WS3 (35). Plasmid pPC106 was linearized with *Hin*dIII and then digested with *Bal* 31, circularized with *Bam*HI linkers (New England BioLabs, Inc., Beverly, Mass.), and ligated to produce plasmid pPC114. Plasmid pPC124 was the 1,977-bp *Hae*III A fragment from plasmid pPC101 ligated to *Hin*cII-linearized plasmid pUC9. Plasmid pPC421 was composed of the insert of plasmid pPC124 cloned in plasmid pUC9 in the opposite orientation. Plasmid pPC145 was constructed by linearizing plasmid pPC124 with *Hin*cII and recircularizing the resultant 4,000-bp restriction fragment. Recombinant phage were plaque purified from a genomic library of strain CS24 constructed with the lambda vector EMBL3 grown on *E. coli* host strains NM538 and NM539 (Promega Biotech, Madison, Wis.).

Colony opacity and M-protein expression determination. Streptococcal colonies were grown and identified as either

* Corresponding author.
† Present address: Department of Biological Sciences, Carnegie-Mellon University, Pittsburgh, PA 15213.

opaque or less opaque phenotypes, as described by Simpson and Cleary (35). M-protein extractions were performed as described by Lancefield and Perlmann (21) and assayed by double diffusion (38). M-protein production by *E. coli* was assayed as described previously (39).

DNA preparation, cloning, and restriction enzyme analysis. Plasmid DNA was prepared from 3- or 500-ml cultures by the alkaline lysis procedure described by Birnboim and Doly (1). Plasmid and phage DNA were digested with restriction enzymes or *Bal* 31 according to the specifications of the manufacturer (New England BioLabs). Restriction enzyme digestion and agarose gel electrophoresis were used to map subclones of plasmids pPC101 and pPC106 and recombinant phage EMBL3-WS3 and to isolate restriction fragments for subsequent subcloning. DNA fragments were electroeluted from agarose onto DE81 paper (Schleicher & Schuell, Inc., Keene, N.H.), eluted from the paper in 20 mM Tris hydrochloride (pH 7.5)–1 mM EDTA–1.5 M NaCl, and ethanol precipitated for further use. Ligations were done as described by Maniatis et al. (23), after linearized vector fragments were dephosphorylated with calf intestinal phosphatase (Boehringer Mannheim Biochemicals, Inc., Indianapolis, Ind.). Rubidium chloride-induced competent *E. coli* host cells were transformed with recombinant plasmids (23).

DNA sequencing. The DNA sequencing strategy used in this study is shown in Fig. 2. Specific restriction fragments of cloned DNA were chemically sequenced by the method described by Maxam and Gilbert (24). Purification of M13 plaques and preparation of single-stranded M13 DNA were performed as described by Messing (25). Overlapping sets of recombinant M13 deletion clones were prepared by the method described by Dale et al. (5) and sequenced by the dideoxy chain-termination method (34) with the universal M13 primer (15).

Streptococcal RNA preparation. Cultures of 100 ml of streptococci were grown to a concentration of approximately 10^8 cells per ml and then centrifuged at 6,000 × *g* for 5 min. The cell pellet was suspended in 20 ml of Todd-Hewitt broth (pH 6.1) supplemented with 30% (wt/vol) sucrose; $MgCl_2$ and dithiothreitol were added to 1 and 0.5 mM, respectively, immediately before protoplasts were prepared. The suspended cells were then incubated with phage lysin (9) at 37°C for 15 min with occasional shaking. Protoplasts were then recovered by centrifugation at 10,000 × *g* for 5 min in an RNase-free Corex tube at 4°C and suspended in 3 ml of ice-cold 4 M guanidinium isothiocyanate (Fluka Chemical, Inc., Haupauge, N.Y.) containing 0.05% Sarkosyl (CIBA-GEIGY Corp., Summit, N.J.), 0.1 M β-mercaptoethanol, and 25 mM sodium citrate. Protoplast disruption was completed by drawing the solution through an 18-gauge needle until a homogeneous solution resulted. CsCl (1.2 g) was then added, and the suspension was vortexed to yield a final volume of 3.5 ml. This solution was layered onto a 1.4-ml 5.7 M CsCl solution containing 100 mM EDTA (pH 7.0) in a polyallomar tube (326819; Beckman Instruments, Inc., Fullerton, Calif.) and then centrifuged in a rotor (SW50.1; Beckman) for 16 h at 21°C. The pelleted RNA was suspended in 450 μl of ice-cold distilled H_2O treated with diethylpyrocarbonate (Sigma Chemical Co., St. Louis, Mo.). The RNA suspension was vortexed and centrifuged in an Eppendorf centrifuge for 1 min at 4°C, and the resultant supernatant ethanol was precipitated with a 1/10 volume of 3 M sodium acetate (pH 5.2) and 2.2 volumes of ethanol at 70°C. The RNA was stored in ethanol or diethylpyrocarbonate-treated H_2O after it was washed in 70% ethanol. Cultures of 100 ml of streptococci averaged a yield of about 2 mg of RNA.

DNA oligomer synthesis. Oligomers were synthesized on a DNA synthesizer (Biosearch) and purified by polyacrylamide gel electrophoresis by the method described by E. Retzel and K. Staskus (personal communication).

Primer extension and RNA sequencing. Total cellular RNAs (16 μg) were hybridized with oligomeric DNA (2 ng) in H_2O by boiling the reactions for 5 min, followed by slow cooling to 42°C. Second-strand synthesis was accomplished by incubating the primer-template complex at 42°C for 15 min with 5 U of avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim) in a reaction mixture of 50 mM Tris hydrochloride (pH 7.8)–40 mM KCl–5 mM $MgCl_2$–10 mM dithiothreitol–100 μM dGTP and dATP–400 μM dTTP–50 μCi of [α-^{32}P]dCTP (3,000 Ci/mM). Reactions were chased with 100 μM dATP, dGTP, dCTP, and dTTP at 37°C for 10 min and terminated by the addition of an equal volume of a mixture containing 98% formamide, 50 mM EDTA (pH 8.0), and 0.1% bromophenol blue and cyanol blue. Samples were then boiled for 3 min, loaded onto an 8% polyacrylamide–7 M urea gel, and electrophoresed (1,000 V for 3 h) to determine the size of the primer extension reaction products. The RNA sequencing reaction mixture was as described above for the primer extension reaction, except that 50 μCi of [α-^{35}S]dATP was substituted for [α-^{32}P]dCTP, and dideoxynucleotide triphosphates were substituted for deoxynucleotides in each of four reactions, as described by Hamlyn et al. (13).

Northern and dot blot analysis. Formaldehyde agarose gel electrophoresis was performed essentially as described by Maniatis et al. (23), except that the final running buffer was 20 mM MOPS (morpholinepropanesulfonic acid), 5 mM sodium acetate, and 1 mM EDTA (pH 7.0); and ethidium bromide was added to the gel at a final concentration of 1 μg/ml. RNA was blotted onto nitrocellulose in 6× SSC buffer (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate; Schleicher and Schuell), as described by Thomas (41). Prehybridization and hybridization with oligomer probes were performed at a T_m − 5°C in 4× SET (1× SET is 25 mM NaCl, 1.5 mM Tris [pH 7.4], and 0.1 mM EDTA) 0.1% sodium PP_i, 0.2% sodium dodecyl sulfate, and 50 μg of heparin per ml, as described by Singh and Jones (36); nick-translated probes (31) were used by the procedure described by Thomas (41). Oligomer probes were 3′ labeled to a specific activity of greater than 10^9 cpm/μg with terminal deoxynucleotidyl transferase (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) by the specifications of Collins and Hunsaker (4).

RESULTS

Deletion analysis. Plasmid pPC106 was described previously (39) as a deletion subclone of pPC101 which encodes a cross-reacting M12 protein. Although these data indicate the direction of the M12 gene transcription and demonstrate that the boundary of the streptococcal insert interrupts the M12 coding sequence, regulatory elements controlling gene expression, including the streptococcal promoter, were not identified. Plasmid DNAs of pPC101 and pPC106 were the source of restriction fragments for construction of recombinant DNA subclones which define the streptococcal DNA required for expression of the M protein gene. The subcloning is described in Fig. 1, and the ability of DNA fragments to express M12 protein is summarized. We found that plasmids pPC113 and pPC124 both express M12 protein,

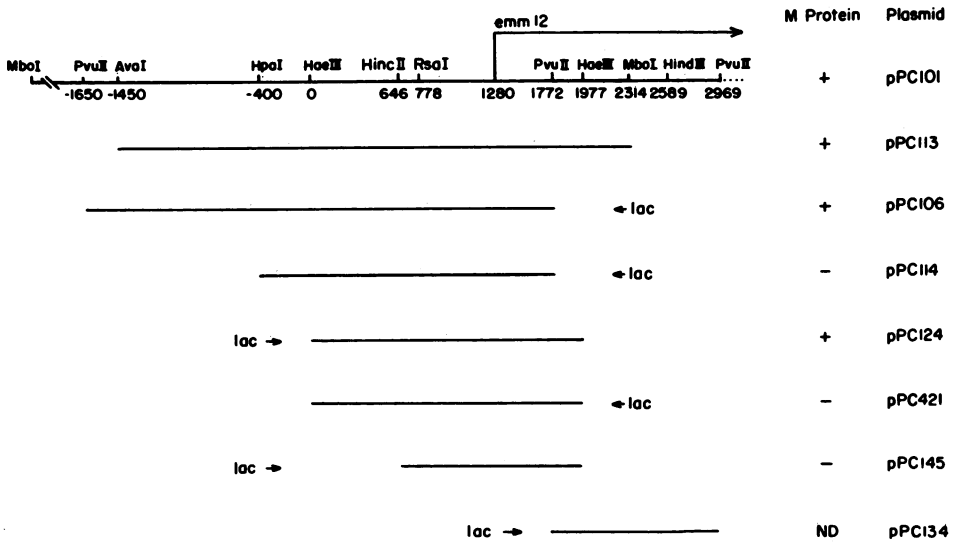

FIG. 1. Restriction endonuclease map of the *Mbo*I-*Pvu*II fragment of *S. pyogenes* DNA encoding the type 12 M protein. The restriction map coordinates indicate the base pairs from the *Hae*III site at base 0. The ability of plasmid subclones to express the M protein that was detectable by double-diffusion analysis is shown next to the indicated restriction fragment cloned in either pUC18 or pUC19. Plasmid constructs are described in the text. ND, Not done.

which is antigenically identical to M protein expressed by pPC101 and that found on the surface of streptococcal strain CS24 when analyzed by immunodiffusion. The elimination of ~1.6 kilobases (kb) of DNA upstream of the M12 coding sequence, a *Pvu*II-*Hae*III fragment (bases −1650 to 0), prevents expression of M protein by plasmid pPC421 in strain JM103. Concordantly, we also observed that M protein is not expressed by plasmids pPC114 and pPC145. Based on data presented later in this report (see Fig. 4), we know that the promoter region controlling M-protein synthesis was not eliminated in these subclones; therefore, we conclude that this upstream region is required for M-protein gene expression. Expression of M protein from the streptococcal insert of plasmid pPC124 may require that it be oriented in the same direction as the *lac* promoter. M protein was not produced by clones that contained this insert in the opposite orientation, i.e., pPC421 (Fig. 1). Thus, expression by pPC124 may be the result of transcription originating from the *lac* promoter. By contrast, plasmid pPC145 did not express M antigen, even though *emm12* was oriented in the same direction as the *lac* promoter. This suggests that the sequence between the 5′ boundaries of the streptococcal insert of pPC124 and pPC145 is required for expression of *emm12* independent of the transcription start site. To better define the boundaries of the *emm12* gene and adjacent regulatory regions, cloned DNA was sequenced and primer extension analysis of streptococcal mRNA was performed.

DNA sequencing of *emm12* and upstream sequences. DNA sequencing of the ~2.0-kb *Hae*III fragment (bases 0 to 1977) was performed to determine the potential coding regions for M protein, the M-protein promoter, and possible upstream regulatory elements. Analysis of 2.97 kb of DNA sequence data (Fig. 2) derived by sequencing portions of plasmids pPC101, pPC106, and pPC134 by the sequencing strategy summarized in Fig. 3 revealed a single open reading frame that could encode type 12 M protein. A predicted amino acid composition of a primarily hydrophilic peptide starting at base 1280 (Fig. 2) was similar to that described previously (40) for type 12 M protein; furthermore, the open reading frame began with a potential hydrophobic signal peptide of 41 amino acid residues, similar to that found in the type 6 M protein (14), and is preceded by a possible ribosome binding site (TAAGGAGC) (37), seven bases upstream of the putative translation start at base 1280 (Fig. 2). An alternative ATG at base 1556 was not considered a likely candidate for a translation start codon of a bacterial surface protein, as neither a signal sequence nor a ribosome binding site was indicated by the DNA sequence. The putative M12 open reading frame encodes a peptide of 73,075 M_r; its 3′ end was also found to be 98% homologous (582 of 598 bp; 2375 to 2972 bp) with the carboxy terminus of the type 6 M protein (Fig. 2). The cloned M12 coding region lacked a termination codon; by analogy with *emm6*, 35 bp of the coding region of *emm12* was undetermined.

Plasmid pPC106 was suggested to encode a truncated M12 protein based on immunodiffusion comparison with extracted streptococcal antigen (39). This conclusion is compatible with the fact that this insert includes nucleotides 1280 to 1772, which would direct the synthesis of an amino-terminal fragment of the M12 antigen. A synthetic peptide corresponding to the first 25 amino acids of the mature M12 protein and predicted from this DNA sequence has been shown to induce antibody which is opsonic for type M12 streptococci (E. Beachey and P. Cleary, unpublished data). Based on these observations, we conclude that the M12 gene (*emm12*) must be encoded by this reading frame and confirm that the plasmid pPC101 insert lacks the carboxy terminus of the M12 protein. Two other reading frames also contain translation start codons that could encode other peptides (base < 0 to 279 and base 299 to 1000). A codon preference analysis (8) of the larger of these two open reading frames suggests that the codon usage is not random, which is a characteristic of open reading frames that encode proteins.

Transcription initiation of the *emm12* gene. A 20-base DNA oligomer that complemented the leader sequence of the presumptive *emm12* gene and no other known *emm12* gene sequence was synthesized, hybridized with strain CS24 total cellular RNA, and used in a primer extension reaction to determine the transcription start of the *emm12* gene. The results suggest that *emm12* in the wild-type strain CS24 is

FIG. 2. DNA sequence of the *Hae*III-*Mbo*I restriction fragment encoding *emm12*. Solid-line and broken-line arrows indicate the transcription start sites of *emm12*; −10 and −35 regions are noted upstream of the transcription start sites. The arrow extending from bases 1309 to 1290 indicates the synthetic oligonucleotide complementary to the DNA sequence used in the primer extension experiments. Translation start codons are noted at bases 299 and 1280, and a stop codon is noted at base 1000 by half boxes. The *emm12* gene in *S. pyogenes* extends beyond the terminal base (2972) noted here. Direct repeats (10 of 11 bases) extending from base 697 to 707 and base 774 to 784 are underlined. Repeat regions A, B, and C are bracketed. The *emm12* homologous carboxy-terminal region is indicated by brackets; bases that differed from the *emm6* gene are underlined. r.b.s., Ribosome-binding site.

transcribed at nearly equal frequencies from two tandem, overlapping promoters separated by 30 bases (Fig. 4a). Sequencing of the RNA (Fig. 4b) demonstrated that each transcription start site consists of a set of three bases, where the center base is the most common start site. The promoter corresponding to the transcription start site producing the longer transcript (P_1) is composed of a −10 region (TATTA TTTAA) that is located within a larger, AT-rich region (78 of 110 bp upstream of the ribosome binding site); the −35 sequence (CTGGTCTTTACC) is in agreement with, at most, four of six consensus sequence bases (TTGACA). Transcription from this promoter is also found to proceed from a purine. The promoter sequence proximal to *emm12* (P_3) differs from the consensus sequences, and its corresponding transcription start site is a pyrimidine base. The −10 region is (CTGAAAAAT), and the −35 region is (TTTACAA TAGA). We also used primer extension analysis to determine the 5′ end of the *emm12* transcript produced by the low-level M-protein-expressing strain CS64 (Fig. 4a), and found that transcription proceeds from a third, less often used transcription start site (P_2). Detection of this transcript in RNA from strain CS24 was variable; therefore, we are unable to relate P_2 to the expression of M12 protein. It should be noted that the *emm12* gene transcription start site lies more than 400 bp downstream from the deletion which eliminated the *Rsa*I site at base 778 (Fig. 1) that is associated with diminished expression of M12 protein in strain CS64 (39; unpublished data).

Determination of the *emm12* gene transcriptional unit size. We also examined by Northern analysis the size of the

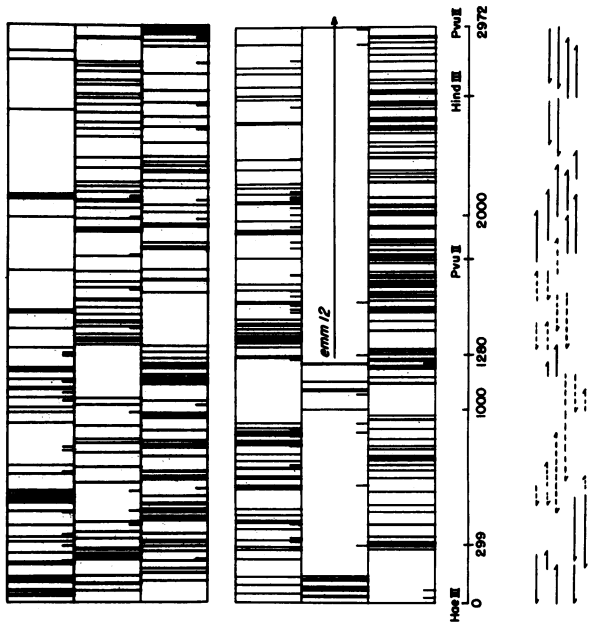

FIG. 3. DNA sequencing strategy and open reading frame map of the 2,969-bp *Hae*III-*Pvu*II restriction fragment encoding *emm12*. Short vertical lines represent potential start codons (AUG), and full-length lines represent termination codons. Each rectangle represents a different reading frame. The solid-line arrows indicate regions sequenced by dideoxy sequencing of M13 subclones of plasmid DNA (34), while broken-line arrows indicate DNA sequenced by the chemical cleavage method described by Maxam and Gilbert (24). The open reading frame encoding *emm12* begins at base 1280 and is indicated by an arrow; an upstream reading frame (base 299 to 1000) is also revealed.

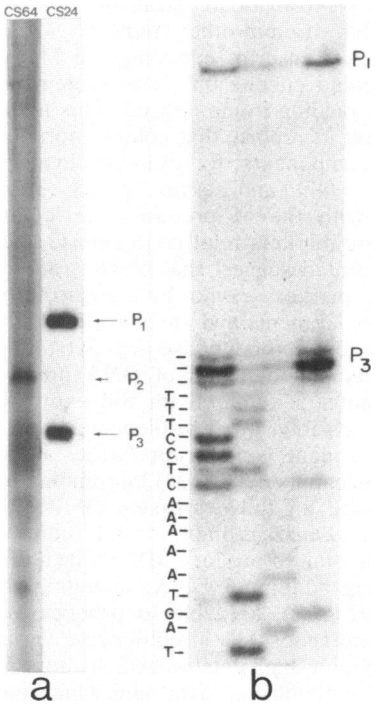

FIG. 4. (a) Primer extension reaction products from M^+ (CS24) and M^- (CS64) strains of *S. pyogenes*. Cellular RNA hybridized with an oligonucleotide complementary to the putative M12 protein-encoding transcript was extended with reverse transcriptase (see text). Reaction products indicating transcription start sites (P_1, P_2, and P_3) were resolved on an 8% acrylamide–7 M urea gel. Molecular weight markers consisted of *Hpa*II-digested pBR322 3′ labeled with [α-^{32}P]dCTP. (b) RNA sequencing ladder of *emm12* transcripts from CS24 with reverse transcriptase. Molecular weight markers consisted of *Hpa*II-digested pBR322 3′ labeled with [α-^{35}S]dCTP.

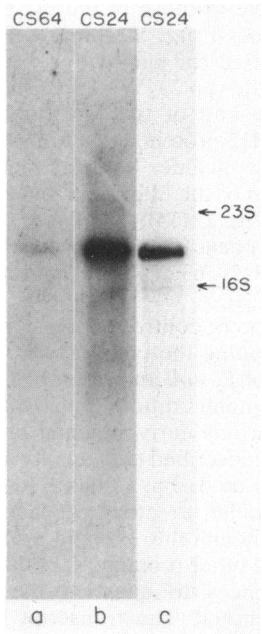

FIG. 5. Northern blot of M^+ (CS24) and M^- (CS64) RNA. Lanes a and b were probed with nick-translated insert DNA from plasmid pPC124. Lane c was probed with a terminal deoxynucleotidyl transferase-labeled oligonucleotide complementary to the *emm12* transcript region encoding the M12 protein leader sequence. *E. coli* rRNA 16S and 23S were used as molecular weight standards.

emm12 transcriptional unit and the extent to which sequences upstream of *emm12* are transcribed. The use of a labeled synthetic oligomer (positions 1290 to 1309) as a probe established that a 2-kb transcript corresponds to *emm12* (Fig. 5, lane c). A transcript of 2 kb can encode one peptide of 60 kilodaltons; therefore, it is likely that *emm12* is monocistronic. Northern blot analysis of strains CS24 and CS64 RNAs (Fig. 5, lanes a and b), with the nick-translated insert from plasmid pPC124 used as a probe (encoding the 5′ end of *emm12* and more than 1 kb of upstream sequence), showed that the diminished M protein in strain CS64 is the consequence of reduced transcription of *emm12* and establishes that a 2-kb transcript corresponds to the *emm12* gene. Minor high-molecular-weight bands (Fig. 5, lane b) may represent unprocessed, full-length transcripts or transcripts emanating from the region adjacent to and upstream of the *emm12* gene. Transcription of *emm12* in strain CS64 is diminished by more than 100-fold as compared with its parent, strain CS24, according to dot blot analysis by using a double-stranded probe (insert DNA from plasmid pPC124) (Fig. 6).

DISCUSSION

The expression of numerous group A streptococcal genes, M proteins (10, 18, 39), streptococcal pyrogenic toxin (16, 45), streptolysin (19), and streptokinase (22) by *E. coli* demonstrates that *E. coli* RNA polymerase is able to recognize promoter sequences of at least some genes from gram-positive organisms. However, little is known about the fidelity of transcription from these promoters or whether controlling elements native to streptococci function in *E. coli*.

The *emm12* gene and the adjacent 5′ DNA previously cloned by our laboratory (39) were sequenced and analyzed in this study to define promoter and other regulatory genes that could participate in the genetic instability of the M^+ phenotype. The DNA sequence data show that a termination codon is not present in the reading frame encoding *emm12*, a fact which indicates that the clones analyzed thus far lack the carboxy terminus of M protein and which predicts that *emm12* encodes a peptide of >73,075 M_r, a size that is greater than those reported for the M12 protein from other strains (58,000 and 64,000 M_rs) (11, 43, 44) but that is consistent with the size variation reported for M protein (11). If the unsequenced end of *emm12* encodes a carboxy terminus identical to that of the M6 protein, then the M12 protein is 74,662 M_r. The predicted protein physically resembles other M proteins of known sequence; first, the first 41 amino acids corresponding to the M12 leader sequence

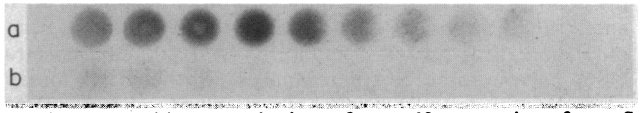

FIG. 6. Dot blot quantitation of *emm12* transcripts from *S. pyogenes*. Duplicate samples of stepped (1:2) dilutions of RNA from M^+ cells, strain CS24 (lane a), and M^- cells, strain CS64 (lane b), were probed with nick-translated insert DNA from plasmid pPC124. Strain CS24 RNA was one-half as concentrated as strain CS64 RNA.

are nearly identical to those of the M1 (26a), M24, and M6 proteins (E. Haanes-Fritz, V. Burdett, E. H. Beachey, and P. Cleary, submitted for publication); second, the carboxy end of the molecule (bp 2375 to 2972) shares 98% homology with the carboxy end of the M6 molecule. The carboxy terminus of the M12 protein constitutes a constant region of the M protein that includes both the membrane anchor and proline-rich region of the M protein, as well as both repeat C (99 bp) regions (Fig. 2) (14). The *emm12* gene also contains two other direct repeat units of greater than 80% homology: repeat A (75 bp; 84%; 63 of 75 bp) and repeat B (75 bp; 88%; 66 of 75 bp) (Fig. 2).

Multiple promoters control expression of numerous bacterial genes, including the *carA* (2, 28), *glnA* (30), and M1 RNA (27) genes of *E. coli* and the *spoVG* gene of *B. subtilis* (17). Tandem promoters differ in their spacing, strength, and activity under various enrivonmental or physiological conditions. Here, we described two equally active promoters, P_1 and P_3, separated by 31 bp. P_1 has −10 and −35 sequences that are very similar to other procaryotic consensus sequences (32). Although the −10 and −35 regions of P_3 were similar to those of other promoters, both varied enough from consensus sequences to question the role of P_3 as an independent promoter. The transcript emanating from P_3 could be an artifact of reverse transcription during the primer extension reaction, as a string of six adenylates immediately downstream of the P_3 start may cause reverse transcription to stall at this point. The possibility that the 5′ termini of the isolated mRNA are the result of processing should also be considered, and the potential of these sequences to bind RNA polymerase should also be tested. Experiments are in progress which will further test the importance of these start sites in the streptococcal cell. Should these transcripts prove to be initiated from separate promoters, we presume that they are differentially expressed under specific environmental conditions and provide the streptococcal cell with the potential to quantitatively vary the amount of M protein on its surface. At this time our information is inadequate to allow speculation as to the role that dual promoters might play in the pathogenesis of this organism.

The DNA sequences directly upstream of both *emm12* (bp 1115 to 1279) and *emm6* are nearly identical (163 of 165 bp), a fact that is consistent with the high degree of homology found to exist in this region by Southern blot analysis (12). Hollingshead et al. (14), in reporting the nucleotide sequence of the *emm6* gene, identified three potential promoters upstream of the *emm6* gene. Our experimentally demonstrated transcription start sites and deduced promoter sequences are in variance with the sequences they designated to be promoters based on comparisons with the −10 and −35 procaryotic promoter consensus sequences (32). Clarification of these differences awaits the experimental demonstration of the *emm6* transcription start sites.

Northern analysis of strain CS64, a spontaneous M^- variant of the M^+ strain CS24, demonstrated a decreased level of M-protein-specific mRNA which accompanied the small deletion of DNA carried by this culture. Primer extension of M-protein-specific mRNA produced by strain CS64 corroborated Northern analysis in that RNA initiated at P_1 and P_3 was absent. A third minor mRNA species, however, also hybridized to the oligomeric probe (Fig. 4a). Our experimental data do not relate this mRNA species to M-protein expression, nor can we be certain that it is not transcribed from a gene which is partially homologous to the probe. Although this third minor species of mRNA is not detected in RNA from the M^+ strain in this study, it has been detected in this strain in other studies.

The primary transcript encoding the M protein is of a length that could encode only one protein, assuming that only a single reading frame is used. This finding has implications in light of reports that colony morphology (35) and cell surface components, for example, hyaluronic acid (3), C5a peptidase (47), and serum opacity (3), appear to be coregulated with the M protein. The location of genes encoding these markers relative to *emm12* is unknown, but the data presented suggest that there is another means by which these markers could be concurrently expressed. Subcloning experiments and analysis of the M^-, hyaluronic acid-negative (3), C5a peptidase-negative (47) mutant strain CS64 suggests that a region of DNA upstream from the *emm12* promoter is required for full expression of the M protein and, possibly, other virulence factors.

Two independent genetic approaches indicate that the DNA sequence outside the M12 promoter and structural gene is required for full expression (35). Subclones which contain both *emm12* promoters and more than 1 kb of upstream DNA fail to produce M12 antigen in *E. coli*, while those carrying additional DNA, including the *Pvu*II site (−1650 bp) (Fig. 1), are able to produce antigen. Strain CS64, a spontaneous M^- streptococcal variant, has been shown to harbor a small deletion which alters the *Rsa*I site at base 778 (39; unpublished data) and which reduces *emm12* transcription by more than 100-fold. Thus, the presence of an intact promoter and structural gene in streptococcal cells or in *E. coli* does not ensure expression of the *emm12* product; therefore, we postulate that this region encodes either a *trans*-active factor or a *cis*-acting DNA sequence that promotes transcription of the *emm12* promoter. The large size of the required segment of DNA, at least 1 kb, is more indicative of a *trans*-active gene product; this is supported by the fact that the deletion carried by strain CS64 is located in an open reading frame that extends from base 299 to 1000 (Fig. 2). By Southern analysis of other independent M^- variant strains, deletions in this upstream region have also been detected (34). The association of deletions with the M^- phenotype is not fortuitous, as 20 wild-type M^+ cultures from isolated colonies of strain CS24 and 5 M^+ cultures of other M12 strains did not reveal deletions in this region (34; unpublished data).

Our data do not implicate deletion formation as the primary source of genetic instability of the M-protein phenotype, nor do we know whether the nearly identical deletions found in strains CS64 and CS46 (39) are the consequence of random mutations or programmed genetic events. We propose that direct repeats (10 of 11 bp) at 697 to 707 bp and 774 to 784 bp in strain CS24 could provide sites for homologous recombination which could result in deletion of 77 bp and the elimination of the *Rsa*I site at position 778.

The requirement of the *arg* locus for the expression of numerous extracellular gene products, including the toxic shock syndrome toxins, in *Staphylococcus aureus* (29) and the coordinate expression of distantly mapped genes encoding virulence factors in *Bordetella pertussis* (46) serve as a model for our interpretation of how expression of multiple streptococcal markers is regulated. Our results prompt us to postulate that streptococci contain a constellation of genes encoding cell surface and extracellular proteins that are controlled by a common regulatory element. The fact that this element is required for the expression of *emm12* in *E. coli* indicates that the fidelity of transcription of these

streptococcal genes is maintained in *E. coli* and will facilitate our future investigation of this regulatory circuit.

ACKNOWLEDGMENTS

This study was supported by Public Health Service grant AI16722 from the National Institute of Allergy and Infectious Diseases. J.C.R. and J.G.S. were supported by Public Health Service training grant 5T32HL107114 from the National Heart and Lung Institute as predoctoral and postdoctoral trainees, respectively.

We thank Michael Williams for helpful discussions.

LITERATURE CITED

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
2. **Bouvier, J., J. Patte, and P. Stragier.** 1984. Multiple regulatory signals in the control region of the *Escherichia coli carAB* operon. Proc. Natl. Acad. Sci. USA **81**:4139–4143.
3. **Cleary, P. P., and Z. Johnson.** 1977. Possible dual function of M protein: resistance to bacteriophage A25 and resistance to phagocytosis by human leukocytes. Infect. Immun. **16**:280–292.
4. **Collins, M. L., and W. R. Hunsaker.** 1985. Improved hybridization assays employing tailed oligonucleotide probes: a direct comparison with 5′ end labeled oligonucleotide probes and nick-translated plasmid probes. Anal. Biochem. **151**:211–224.
5. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13**:31–40.
6. **Eisenstein, B. I.** 1981. Phase variation of type 1 fimbriae in *Escherichia coli* is under transcriptional control. Science **214**: 337–338.
7. **Eisenstein, B. I.** 1982. Operon fusion of the phase variation switch. A virulence factor in *Escherichia coli*. Infection **10**: 112–115.
8. **Fickett, J.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. **10**:5303–5318.
9. **Fischetti, V. A., E. C. Gotschlich, and A. W. Bernheimer.** 1971. Purification and physical properties of group C streptococcal phage-associated lysin. J. Exp. Med. **133**:1105–1117.
10. **Fischetti, V. A., K. F. Jones, B. N. Manjula, and J. R. Scott.** 1984. Streptococcal M6 protein expressed in *Escherichia coli*. Localization, purification, and comparison with streptococcal-derived M protein. J. Exp. Med. **159**:1083–1095.
11. **Fischetti, V. A., K. F. Jones, and J. R. Scott.** 1985. Size variation of the M protein in group A streptococci. J. Exp. Med. **161**: 1384–1401.
12. **Haanes-Fritz, E., J. C. Robbins, and P. Cleary.** 1987. Comparison of genes encoding group A streptococcal M protein types 1 and 12: conservation of upstream sequences, p. 106–108. *In* J. Ferretti and R. Curtiss III (ed.), Streptococcal genetics. American Society for Microbiology, Washington, D.C.
13. **Hamlyn, P. H., G. G. Brownlee, C. Cheng, M. J. Gait, and C. Milstein.** 1978. Complete sequence of constant and 3′ noncoding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing. Cell **15**:1067–1075.
14. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1986. Complete nucleotide sequence of type 6 M protein of the group A streptococcus. Repetitive structure and membrane anchor. J. Biol. Chem. **261**:1677–1686.
15. **Hu, N., and J. Messing.** 1982. The making of strand-specific M13 probes. Gene **17**:271–277.
16. **Johnson, L. P., and P. M. Schlievert.** 1984. Group A streptococcal phage T12 carries the structural gene for pyrogenic exotoxin type A. Mol. Gen. Genet. **194**:52–56.
17. **Johnson, W. C., C. P. Moran, and R. Losick.** 1983. Two RNA polymerase sigma factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. Nature (London) **302**:800–804.
18. **Kehoe, M. A., T. P. Poirier, E. H. Beachey, and K. N. Timmis.** 1985. Cloning and genetic analysis of serotype 5 M protein determinant of group A streptococci: evidence for multiple copies of the M5 determinant in the *Streptococcus pyogenes* genome. Infect. Immun. **48**:190–197.
19. **Kehoe, M., and K. N. Timmis.** 1984. Cloning and expression in *Escherichia coli* of the streptolysin O determinant from *Streptococcus pyogenes*: characterization of the cloned streptolysin O determinant and demonstration of the absence of substantial homology with determinants of other thiol-activated toxins. Infect. Immun. **43**:804–810.
20. **Lancefield, R. C.** 1959. Persistence of type-specific antibodies in man following infection with group A streptococci. J. Exp. Med. **110**:271–292.
21. **Lancefield, R. C., and G. E. Perlmann.** 1952. Preparation and properties of type-specific M antigen isolated from a group A, type 1 hemolytic streptococcus. J. Exp. Med. **96**:71–82.
22. **Malke, H., and J. J. Ferretti.** 1984. Streptokinase: cloning, expression, and excretion by *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81**:3557–3561.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook (ed.).** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
25. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
26. **Meyer, T. F., N. Mlower, and M. So.** 1982. Pilus expression in *Neisseria gonorrhoeae* involves chromosome rearrangement. Cell **30**:45–52.

26a. **Morávek, L., O. Kühnemund, J. Havliček, P. Kopecký, and M. Pavlik.** 1986. Type 1 M protein of *Streptococcus Pyogenes*: N-terminal sequence and peptic fragments. FEBS Lett. **208**: 435–438.

27. **Motemedi, H., Y. Lee, and F. J. Schmidt.** 1984. Tandem promoters preceding the gene for the M1 RNA component of *Escherichia coli* ribonuclease P. Proc. Natl. Acad. Sci. USA **81**:3959–3963.
28. **Piette, J., H. Nyunoya, C. J. Lusty, R. Cunin, G. Weyens, M. Crabeel, D. Charlier, N. Glansdorff, and A. Pierard.** 1984. DNA sequence of the carA gene and the control region of *carAB*: tandem promoters, respectively controlled by arginine and the pyrimidines, regulate the synthesis of carbamoylphosphate synthetase in *Escherichia coli* K 12. Proc. Natl. Acad. Sci. USA **81**:4134–4138.
29. **Recsei, P., B. Kreiswirth, M. O'Reilly, P. Schlievert, A. Gruss, and R. P. Novick.** 1986. Regulation of exotoxin gene expression in *Staphylococcus aureus* by *agr*. Mol. Gen. Genet. **202**:58–61.
30. **Reitzer, L. J., and B. Magasanik.** 1985. Expression of *glnA* in *Escherichia coli* is regulated at tandem promoters. Proc. Natl. Acad. Sci. USA **82**:1979–1983.
31. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
32. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
33. **Rothbard, S., and R. F. Watson.** 1948. Variation occurring in group A streptococci during human infection. Progressive loss of M substance correlated with increasing susceptibility to bacteriostasis. J. Exp. Med. **87**:521–535.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Simpson, W. J., and P. P. Cleary.** 1987. Expression of M type 12 protein by a group A streptococcus exhibits phase like variation: evidence of coregulation of colony opacity determinants and M protein. Infect. Immun. **55**:2448–2455.
36. **Singh, L., and K. W. Jones.** 1984. The use of heparin as a cost-effective means of controlling background in nucleic acid hybridization procedures. Nucleic Acids Res. **12**:5627–5638.
37. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci.

USA **71**:1342–1346.
38. **Spanier, J. G., and P. P. Cleary.** 1980. Bacteriophage control of antiphagocytic determinants in group A streptococci. J. Exp. Med. **152**:1393–1406.
39. **Spanier, J. G., S. J. C. Jones, and P. Cleary.** 1984. Small DNA deletions creating avirulence in *Streptococcus pyogenes*. Science **225**:935–938.
40. **Straus, D. C., and C. F. Lange.** 1972. Immunochemistry and end-group analysis of group A streptococcal M proteins. Infect. Immun. **5**:927–932.
41. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
42. **Todd, E. W., and R. C. Lancefield.** 1928. Variant of hemolytic streptococci; their relation to type-specific substance, virulence, and toxin. J. Exp. Med. **48**:751–767.
43. **van de Rijn, I., and V. A. Fischetti.** 1981. Immunochemical analysis of intact M protein secreted from cell wall-less streptococci. Infect. Immun. **32**:86–91.
44. **Vosti, K. L., R. H. Johnson, and M. F. Dillon.** 1971. Further characterization of purified fractions of M protein from a strain of group A, type 12 streptococcus. J. Immunol. **107**:104–114.
45. **Weeks, C. R., and J. J. Ferretti.** 1986. Nucleotide sequence of the type A streptococcal exotoxin (erythrogenic toxin) gene from *Streptococcus pyogenes* bacteriophage T12. Infect. Immun. **52**:144–150.
46. **Weiss, A. A., and S. Falkow.** 1984. Genetic analysis of phase change in *Bordetella pertussis*. Infect. Immun. **43**:263–269.
47. **Wexler, D. E., R. D. Nelson, and P. P. Cleary.** 1983. Human neutrophil chemotactic response to group A streptococci: bacteria-mediated interference with complement-derived chemotactic factors. Infect. Immun. **39**:239–246.